Newsletter of Centre on the Impact of AI and Robotics
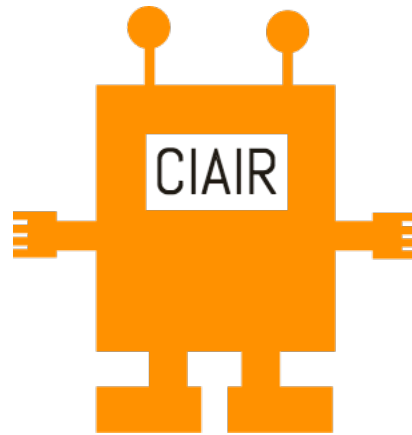
December 2016

Contents

Overview of CIAIR

The Centre's mission is to study the potential impact of AI and robotics, and to promote beneficial outcomes by means of research, teaching, measurement and public debate. Its goal is to drive AI and robotics towards secure, safe and successful returns by developing and influencing thought leaders of tomorrow through an extensive programme of research, education, conferences, workshops and events.

The Centre is highly multidisciplinary, bringing in members from Computer Science, Economics, History, Law, Philosophy, Sociology, and elsewhere, as well as hosting sabbatical and other long stay visitors. The Centre has a Scientific Advisory Board that draws in thought leaders from around the world.

CIAIR is pronounced "Sea Air".

Find us at http://ciair.org

Upcoming CIAIR Events

On November 30th, there will be a Centre meeting to discuss the proposed kickoff projects. More details about these can be found overleaf.

On December 1st, Toby Walsh will be speaking about what in AI keeps him up at night. He will be joining 11 other experts in a public event that launches UNSW's Grand Challenges programme. The event will "blend expertise and entertainment". For more details, see the unsomnia website

From Dec 12th to 16th, the Fifth Review Conference of the Convention on Certain Conventional Weapons will take place in the UN in Geneva. On Dec 15th, Toby Walsh has been invited by Human Rights Watch to speak about the challenges posed by AI in the battlefield.

Finally, in Feb 2017, the Centre will be hosting the 3rd International Workshop on AI, Ethics and Society. It will be co-located with AAAI 2017 in San Francisco. More details here

Government Reports

It's like buses. You wait for government to address concerns about where AI is taking society, and then three reports from government come along in quick succession.

In October, following on from a series of workshops, the White House published a report in October giving recommendations for the Future of AI. It recommended goverment should prioritize basic and long-term AI research, consider effects of AI on jobs, and on warfare. Unfortunately, given the recent US elections, it is unclear if any of this will make it into policy. More details here.

On the same day, the UK parliament issued a report on the impact of AI and robotics. This was less satisfactory than the White House Report. It did not recommend any extra funding for AI, offered no recommendations on autonomous weapons, or on many of the other good ideas in the US report like opening out government to AI based improvements. More details here.

Finally, in November, the UK Government Office for Science issued a report on "Artificial intelligence: opportunities and implications for the future of decision making". More details here.

Related Research Centres

Seven institutes have been set up in the last few years to look at the impact on AI.

Machine Intelligence Research Institute (Berkeley). MIRI was set up in 2000 as the Singularity Institute for Artificial Intelligence. Its goal is to help humanity prepare for the moment when machine intelligence exceeds human intelligence.

Future of Life Institute (Boston). Set up by Max Tegmark (MIT) in 2014. FLI is focused on existential risks facing mankind, many of which come from other directions than AI.

Leverhulme Centre for the Future of Intelligence (University of Cambridge). Set up in 2015 by Huw Price with a £10 million grant. The goal of the Centre is to explore the opportunities and challenges of AI, both short and long term. The Centre will bring together computer scientists, philosophers, social scientists and others to examine the technical, practical and philosophical questions AI raises for humanity in the coming century. The Leverhulme Centre is the closest in focus to CIAIR. Huw Price sits on the Scientific Advisory board of CIAIR, and discussions are taking place to co-host visitors.

Strategic Artificial Intelligence Research Centre (University of Oxford). Set up in 2015 by Nick Bostrom and others. This Centre aims to develop policies for government, industry, and others to minimize risks and maximize benefit from AI over the longer term.

Centre for Human-Compatible AI (University of California at Berkeley). Set up in August 2016 by Stuart Russell with a $5.5 million grant. The Centre is focused on research into AI safety.

Centre on Artificial Intelligence for Society (University of Southern California). Set up in October 2016 by Milind Tambe, a pioneer in the area of security games. This Centre is focused on undertaking research into AI with societal benefit.

Centre for Ethics and Computational Technologies (Carnegie Mellon University). Set up in November 2016 with a $10 million grant. This Centre is focused on on ethics.

Except for the Leverhulme Centre in Cambridge, each of these centres is specialized, focused on one area: existential risk (MIRI, FLI, CHCAI), ethics (CECT), societal benefits (CAIS), and policy (SAIRC). The Leverhulme Centre and CIAIR are the only institutes tasked to consider the broad impact of AI and robotics, not focused just on existential risks, tackling both short and long term issues, and considering a wide range of topics including economical, societal, and legal impacts.

CIAIR Projects under Development

AI AND WORK. Kevin Fox (Economics); Steve Frenkel (Sociology/Organisational psychology); Juan Carlos Carbajal (Economics); Sunghoon Kim (Employment Relations/Human Resource Management); and Amirali Minbashian (Organisational Psychology).

In what sectors of the economy are AI and Robotics being introduced in Australia? To what extent is this welcomed or opposed? Is this related to particular (mis) conceptions of these new technologies? How can we characterize and model the adoption, use and effectiveness of AI and Robotics? What are the sources and mechanisms that explain this? What are the short and longer term consequences of the diffusion of AI and Robotics? How does the Australian experience compare with other countries? How can similarities and differences be explained, and what lessons can be learnt from these comparisons?

AI AND WAR. Toby Walsh (CSE); Jai Galliott (ADFA).

What are the moral arguments for and against autonomous weapons? Does international law need to be adjusted to deal with autonomous weapons? If so, how? How do we define concepts at the centre of UN discussion like "meaningful human control" and "levels of autonomy"? Should we differentiate between offensive and defensive weapons? How can we support technically the discussions at the UN and elsewhere surrounding autonomous weapons? More generally, how do we ensure robots behave ethically? What ethical laws should such robots follows? Can a robot learn our ethical values? How do we ensure that they cannot be hacked to behave in undesirable ways? How do we promote less controversial military applications for AI like mine clearing? And can we develop a professional code of ethics for researchers working on such systems?

AI AND EDUCATION. Kalervo N. Gulson (Arts & Social Sciences), Matthew Kearnes (Arts & Social Sciences), Edward Scheer (Arts & Social Sciences), Iain Skinner (Engineering), Andrew Murphie (Arts & Social Sciences).

Key questions to be asked include (1) what are the new ways of learning based upon ML? How do we educate an AI system? How is its learning different?; (2) what are the possibilities and challenges for education and education policy from the use of AI in education? How will robotics (as "actors") and AI (as "authors"), and the connection to big data, change what we understand as representation in the creative arts?; (3) what are the ethical, economical and biosocial considerations of implementing AI in education; and (4) how does ML use ideas from social policy, including policy and value networks, and how can policy analysts use these same ideas?